

NASA-25066

CLOUD FIELD CLASSIFICATION BASED ON TEXTURAL FEATURES

BY

Sailes Kumar Sengupta

Professor of Mathematics and Computer Science

S.D School of Mines & Technology

Rapid City, SD 57701

Prepared for the NASA/ASEE Summer Faculty Program

July 27, 1989

ABSTRACT

An essential component in global climate research is accurate cloud cover and type determination. The type classification can be done by multispectral/textural methods. Of the two approaches to texture-based classification (statistical and textural), only the former is effective in the classification of natural scenes such as, land, ocean, and atmosphere. The reason lies in the fact that in the latter, visual scenes are analyzed in terms of the organization and relationships among its substructures. These aspects are transparent in man-made objects, but not so in natural scenes. In contrast, in the statistical approach that we have adopted, parameters characterizing the stochastic properties of the spatial distribution of grey levels in an image are estimated and then used as features for cloud classification.

Two types of textural measures have been used in this study. One is based on the distribution of the grey level difference vector (GLDV), and the other on a new set of textural features derived from the MaxMin cooccurrence matrix (MMCM).

The GLDV method looks at the difference D of grey levels at pixels separated by a horizontal distance (assuming isotropy) d and computes several statistics based on this distribution. These are then used as features in subsequent classification.

The MaxMin textural features on the other hand are based on the MMCM, a matrix whose (I,J) th entry gives the relative frequency of occurrences of the grey level pair (I,J) that are consecutive and thresholded local extremes (see fig.1) separated by a given pixel distance d . Textural measures are then computed based on this matrix in much the same manner as is done in texture computation using the grey level cooccurrence matrix.

The database consists of 37 cloud field scenes from Landsat imagery using a near IR visible channel. Among these there are 15 stratocumulus, 10 cumulus, and 12 cirrus scenes. Each scene has been subdivided into 20 subregions with each subregion representing a sampling unit. Approximately one third of the scenes chosen at random from each type has been set aside for testing the classification accuracy using a Jackknife validation procedure. The subregions in the remaining scenes of each type have been used for training the two classifiers. The classifiers are based on features computed from the GLDV distribution and from the MMCM distribution respectively. The texture features for each sampling unit have been computed for different horizontal separations d ($d = 1, 2, \dots$ pixels). The operation has been repeated at different levels of resolution of the scenes, obtained by progressive spatial averaging of 2×2 , 4×4 , \dots pixels. Features have also been combined by using combinations of the values of d (such as, $d=1, 2$; $d=1, 4$; $d=1, 2, 3$; \dots etc.) providing more features available for use by the classifiers.

The classification algorithm used in this study is the well known Stepwise Discriminant Analysis. This combines the two steps in a classification procedure, feature selection and construction of the linear discriminant functions, into one. The algorithm, although suboptimal, is known to have worked well in various applications. We have implemented this by the BMDP(TM) package using the procedure BMDP7M.

The overall accuracy has been estimated by the percentage of correct classification in each case.

It turns out that both types of classifiers, at their best combination of features, and at any given spatial resolution, give approximately the same classification accuracy, viz. between 84 to 90 per cent. However, at lower spatial resolution, the best performance appears to be slightly better in the case of the MMCM based classifier. (See figs. 2 and 3)

In an ongoing work, we are using a Neural Network based classifier with a feed forward architecture and a back propagation training algorithm to increase the classification accuracy, using these two classes of features. Preliminary results based on GLDV textural features alone look promising in two ways. First, the classification accuracy has gone up to 93%. Second, the percentage of sampling units needed for training the classifier to attain this level of accuracy is only 20%, compared to 67% of the training samples needed when we use the GLDV based features in our classifier.

MAX-MIN THRESHOLDS

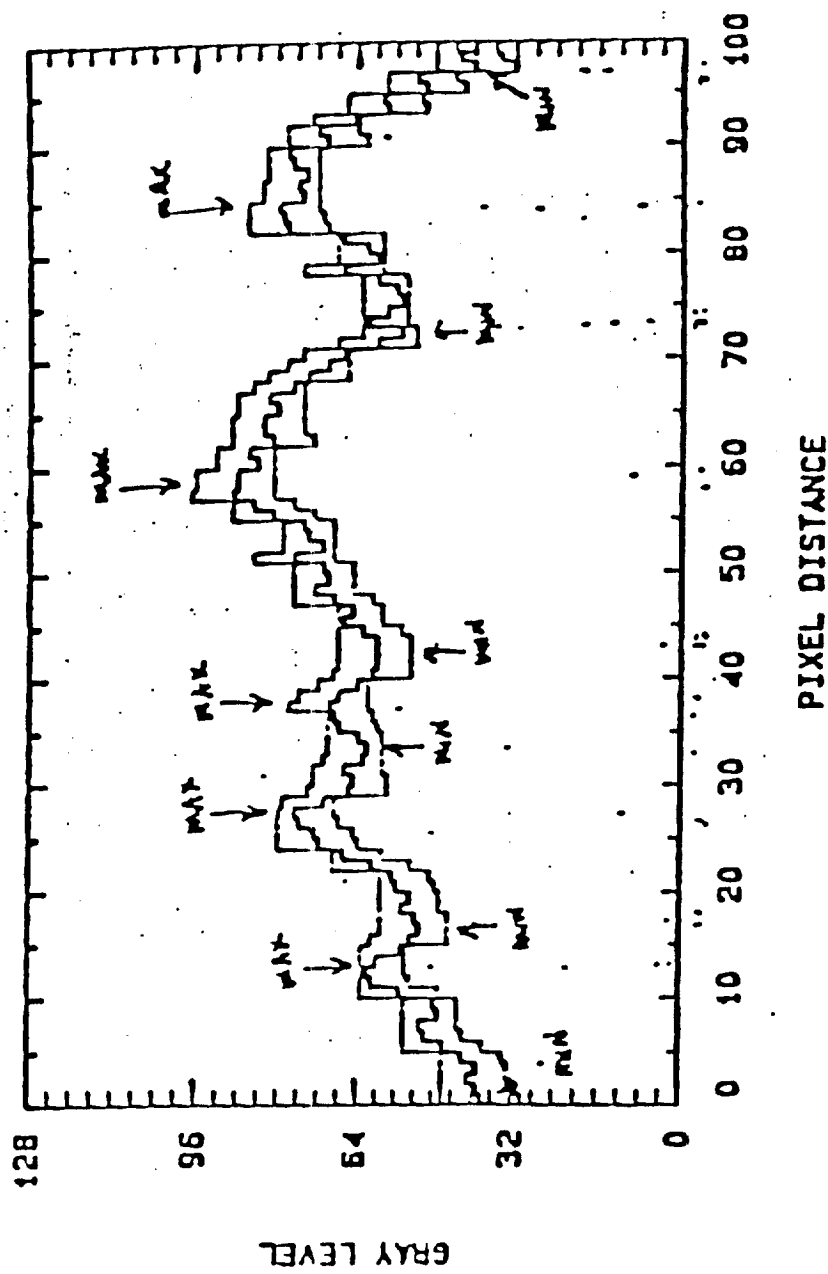
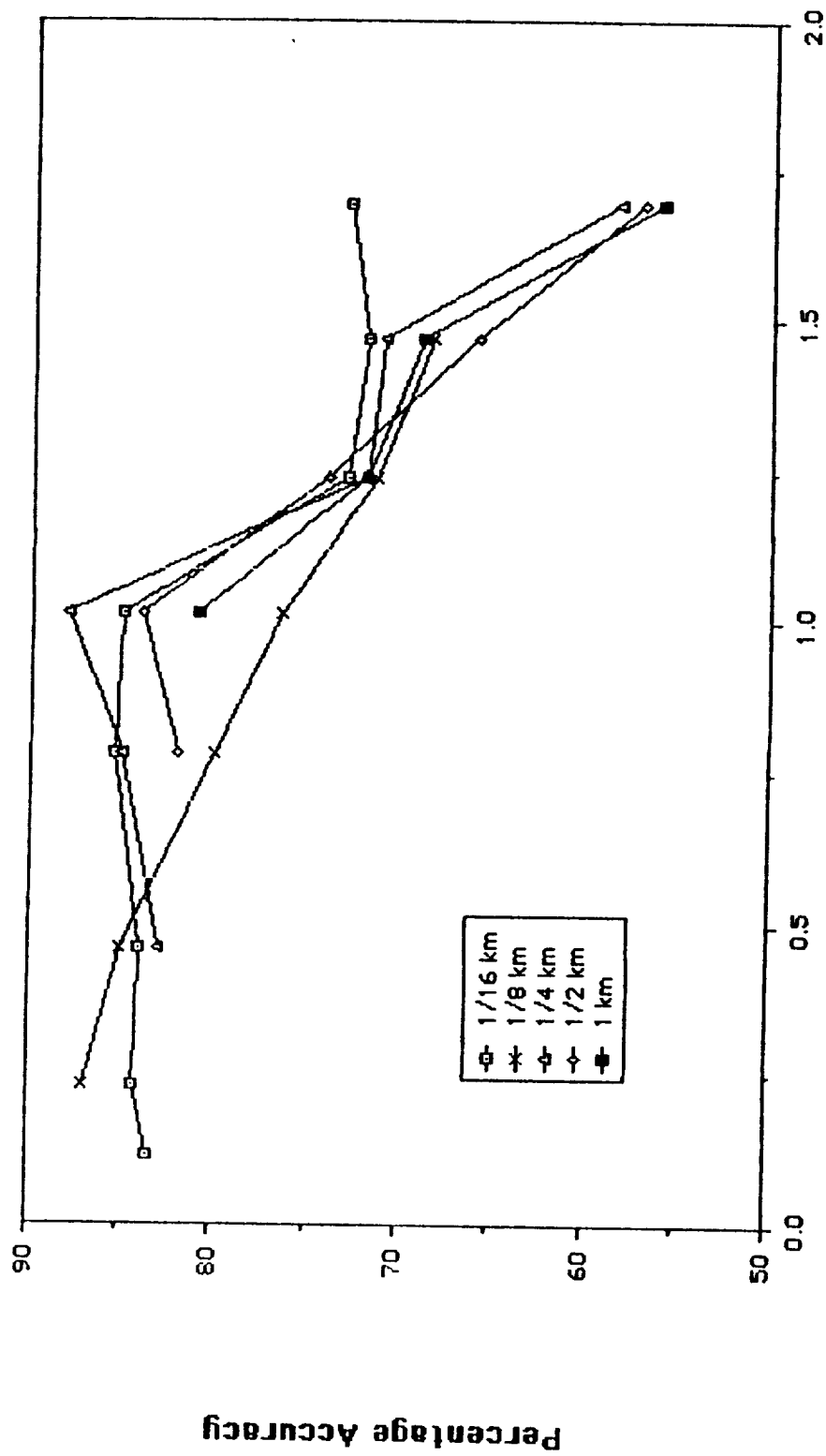


Fig. 1: Operation of the MAX-MIN algorithm for a representative cumulus cloud segment.

Classification Accuracy Plot

Maxmin Cocurrence Texture



Distance (km)

FIG 2

CLASSIFICATION ACCURACY
USING GLDV TEXTURE

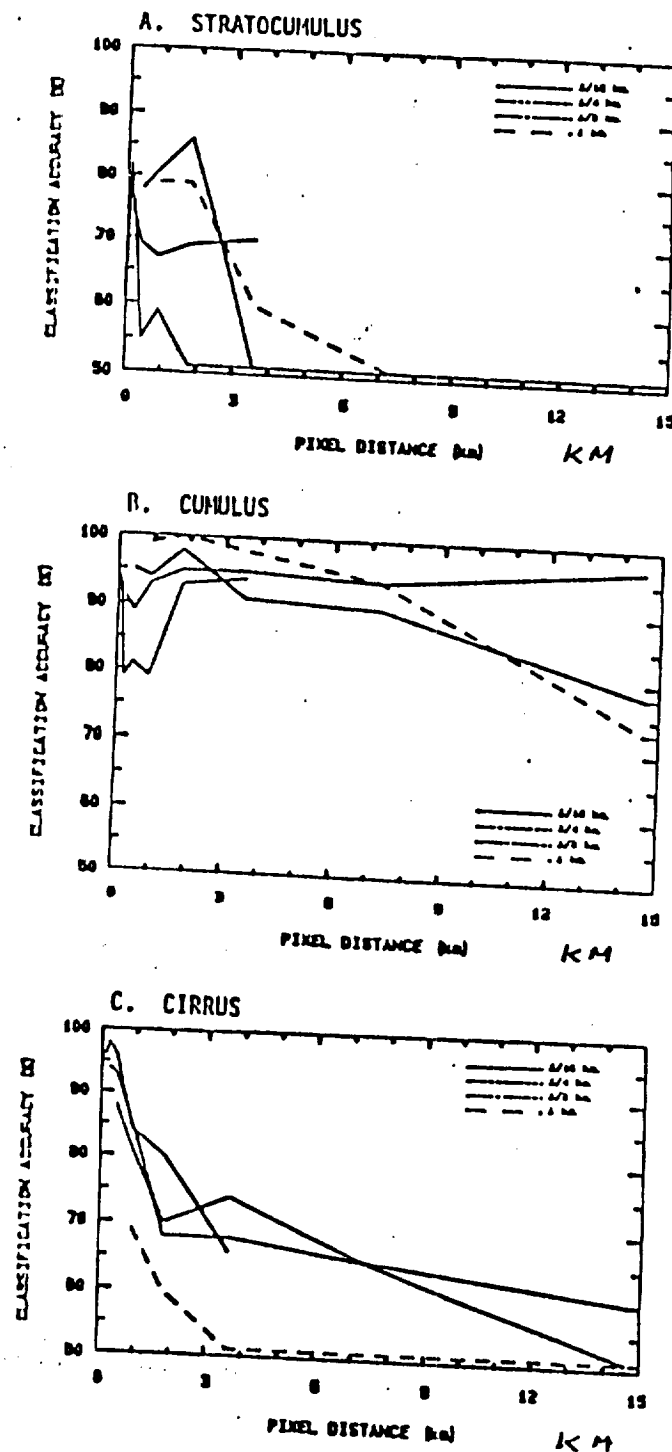


Fig. 3: GLDV classification accuracy for strato-cumulus, cumulus, and cirrus as a function of pixel separation and spatial resolution.